

38


**Europäisches Patentamt**  
**European Patent Office**  
**Office européen des brevets**

Publication number:

**0 079 739**  
**A2**

12

# **EUROPEAN PATENT APPLICATION**

Application number: 82305926.6

Date of filing: 08.11.82

Int. Cl.<sup>3</sup>: **C 12 N 15/00**  
**C 12 N 1/00, C 12 P 21/02**  
**C 07 H 21/04, C 07 C 103/52**  
**//C12R1/19, C12R1/865**

Priority: 12.11.81 US 320632

Date of publication of application:  
25.05.83 Bulletin 83/21

Designated Contracting States:  
BE CH DE FR GB IT LI NL SE

Applicant: **THE UPJOHN COMPANY**  
 301 Henrietta Street  
 Kalamazoo, Michigan 49001(US)

Inventor: **Dugaiczky, Achilles**  
 c/o The Upjohn Company 301 Henrietta Street  
 Kalamazoo Michigan 49001(US)

Representative: **Perry, Robert Edward et al,**  
**GILL JENNINGS & EVERY 53-64 Chancery Lane**  
 London WC2A 1HN(GB)

Albumin-based nucleotides, their replication and use, and plasmids for use therein.

The DNA sequence coding for human serum albumin has been isolated and inserted as two fragments into two novel plasmids which can be replicated in *E. coli*. These novel fragments can be joined to provide a unitary DNA sequence which then can be cloned into a suitable host, e.g. *E. coli*, for the expression of human serum albumin (which is used extensively in medical practice in treating shock conditions).

EP 0 079 739 A2

**BEST AVAILABLE COPY**

ALBUMIN-BASED NUCLEOTIDES, THEIR REPLICATION  
AND USE, AND PLASMIDS FOR USE THEREIN

This invention relates to nucleotides related to human serum albumin (HSA), their replication and use, and plasmids (and host substances) for use therein.

The gene for serum albumin is regulated in development. On the other hand, serum albumin is synthesised in mammals by the adult liver, and its plateau in adulthood. The embryonic liver and yolk sac, on the other hand, produce predominantly  $\alpha$ -fetoprotein, but the synthesis decreases drastically after birth. Recently, Law et al determined the complete sequence of mouse  $\alpha$ -fetoprotein mRNA, Nature 291 (1981) 201-205. The structure revealed extensive homology to mammalian serum albumin, indicating that the two proteins are encoded in the same gene family. Similar conclusions have been reached from studies on the  $\alpha$ -fetoprotein genes of the rat and the mouse; see Jagodzinski et al, Proc. Natl. Acad. Sci. USA, 78 (1981) 3521-3525, and Gorin et al, J. Biol. Chem. 256 (1981) 1954-1959.

The complete nucleotide sequence of human serum mRNA has been determined from recombinant cDNA clones and from a primer-extended cDNA synthesis on the mRNA template. The sequence comprises 2,078 nucleotides, starting upstream of a potential ribosome binding site in the 5'-untranslated region. It contains all the translated codons and extends into the poly(A) at the 3'-terminus. Part of the translated sequence codes for a hydrophobic prepeptide met-lys-trp-val-thr-phe-ile-ser-leu-leu-phe-leu-phe-ser-ser-ala-tyr-ser, followed by a basic propeptide arg-gly-val-phe-arg-arg. These signal peptides are absent from mature serum albumin and, so far, have not been identified in their nascent state in humans. A remaining 1,755 nucleotides of the translated mRNA sequence code for 585 amino acids which are in agreement, with few exceptions, with the published amino acid data for human serum albumin. The mRNA sequence verifies and refines the repeating homology in the triple-domain structure of the serum albumin molecule.

DETAILED DESCRIPTION OF THE INVENTION

Human serum albumin cDNA is cloned into the PstI site of plasmid pBR322 by the oligo(dG)-oligo(dC) tailing technique. Plasmid DNA was isolated from 97 positive colonies which hybridized to the enriched  
 5 albumin cDNA probe, and the recombinant plasmid pHA36 was found to contain the largest insert of an albumin cDNA sequence. Its restriction endonuclease map is shown in the drawing, together with a restriction map of the primer-extended plasmid clone pHA206. The latter was obtained in a second transformation experiment after initiating  
 10 the cDNA synthesis from an internal primer. This primer was a 91 base pairs long DNA fragment, MspI(152)-TaqI(182/3), isolated from pHA36. The two plasmids, pHA36 and pHA206, share 0.15 kb of homologous DNA. Together, they encode the entire sequence for human serum albumin, starting with the CTT codon for leu -10 of the prepeptide and extending  
 15 into the 3'-untranslated region of poly(A).

Sequence of the Albumin cDNA. The sequence was determined for the most part on both DNA strands to ensure accuracy. All of the restriction sites used to end-label DNA fragments were sequenced across by  
 20 labeling a neighboring restriction site. The entire nucleotide sequence of the serum albumin mRNA, as determined from the cloned DNA in pHA36, pHA206, and from the primer-extended cDNA at the 5'-terminus of the message, is shown in the following Table 1. The inferred amino acid sequence is also indicated. The mRNA length is 2,078 nucleotides, of which 38 represent the 5'-untranslated region, 54 identify a  
 25 prepeptide of 18 amino acids, 18 identify a propeptide of 6 amino acids, 1,755 code for the known 585 amino acids of serum albumin, 189 make up the 3'-untranslated region and 24 are the poly(A) sequence. Nucleotides 5 to 15 (-34 to -24) in the 5'-untranslated region (Table  
 30 1) are complementary to a 3'-terminal region of eukaryotic 18S RNA [Azad, A.A. and Deacon, N.J. (1980) Nucl. Acids Res. 8, 4365-4376] and thus could represent a ribosome binding site:

(5')...T T<sup>C</sup>T C T T C T G T.....albumin mRNA  
 35 (3')...G A G G A A G G C G U C C m<sub>2</sub><sup>6</sup>A m<sub>2</sub><sup>6</sup>A.....18S RNA

The translated portion of the mRNA sequence codes for the signal peptide and the main body of the albumin polypeptide chain. The

signal peptide is composed of a hydrophobic prepeptide of 18 amino acids and a basic propeptide of 6 amino acids (Table 1). Since pre-peptides are removed from nascent secretory proteins (like albumin) in the endoplasmic reticulum, they are seen only in vitro in heterologous  
5 translation systems. As yet, they have not been found within cells [Judah, J.D. and Quinn, P.S. (1977) FEBS 11th Mtg., Copenhagen 50, 21-29; and Strauss, A.W., Donohue, A.M., Bennett, C.D., Rodkey, J.A. and Alberts, A.W. (1977) Proc. Natl. Acad. Sci. USA 74, 1358-1362]. This is the first report of the presence and the sequence of a pre-  
10 peptide for human serum albumin. As it is with other secretory proteins, the conversion of proalbumin to albumin takes place in the Golgi vesicles, and the enzyme responsible for this cleavage is probably cathepsin B [Judah, J.D. and Quinn, P.S. (1978) Nature 271, 384-385]. This is also a first report on the sequence of the pro-  
15 peptide for normal human serum albumin.

At the 3'-end of the message, the putative polyadenylation signal sequence, AATAAA, is located 164 nucleotides downstream from the amino acid termination codon TAA and 16 nucleotides upstream from the beginning of the poly(A) sequence. Another characteristic sequence  
20 located near the polyadenylation site has been identified by Benoist, et al. [Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. 8, 127-142]; the consensus sequence from several mRNAs was concluded as TTTCCTCTGC. A similar sequence, TTTCTCTGT, is located 19 nucleotides upstream from the AATAAA hexanucleotide in the  
25 human albumin mRNA (Table 1).

TABLE 1

35	51	81	111	141	171	201	230	260	290	320	350	380	410	440	470	500	530	560	590	620	650	680	710	740	770	800	830	860	890	920	950	980	1010	1040	1070	1100	1130	1160	1190	1220	1250	1280	1310	1340	1370	1400	1430	1460	1490	1520	1550	1580	1610	1640	1670	1700	1730	1760	1790	1820	1850	1880	1910	1940	1970	2000	2030	2060	2090	2120	2150	2180	2210	2240	2270	2300	2330	2360	2390	2420	2450	2480	2510	2540	2570	2600	2630	2660	2690	2720	2750	2780	2810	2840	2870	2900	2930	2960	2990	3020	3050	3080	3110	3140	3170	3200	3230	3260	3290	3320	3350	3380	3410	3440	3470	3500	3530	3560	3590	3620	3650	3680	3710	3740	3770	3800	3830	3860	3890	3920	3950	3980	4010	4040	4070	4100	4130	4160	4190	4220	4250	4280	4310	4340	4370	4400	4430	4460	4490	4520	4550	4580	4610	4640	4670	4700	4730	4760	4790	4820	4850	4880	4910	4940	4970	5000	5030	5060	5090	5120	5150	5180	5210	5240	5270	5300	5330	5360	5390	5420	5450	5480	5510	5540	5570	5600	5630	5660	5690	5720	5750	5780	5810	5840	5870	5900	5930	5960	5990	6020	6050	6080	6110	6140	6170	6200	6230	6260	6290	6320	6350	6380	6410	6440	6470	6500	6530	6560	6590	6620	6650	6680	6710	6740	6770	6800	6830	6860	6890	6920	6950	6980	7010	7040	7070	7100	7130	7160	7190	7220	7250	7280	7310	7340	7370	7400	7430	7460	7490	7520	7550	7580	7610	7640	7670	7700	7730	7760	7790	7820	7850	7880	7910	7940	7970	8000	8030	8060	8090	8120	8150	8180	8210	8240	8270	8300	8330	8360	8390	8420	8450	8480	8510	8540	8570	8600	8630	8660	8690	8720	8750	8780	8810	8840	8870	8900	8930	8960	8990	9020	9050	9080	9110	9140	9170	9200	9230	9260	9290	9320	9350	9380	9410	9440	9470	9500	9530	9560	9590	9620	9650	9680	9710	9740	9770	9800	9830	9860	9890	9920	9950	9980	10010	10040	10070	10100	10130	10160	10190	10220	10250	10280	10310	10340	10370	10400	10430	10460	10490	10520	10550	10580	10610	10640	10670	10700	10730	10760	10790	10820	10850	10880	10910	10940	10970	11000	11030	11060	11090	11120	11150	11180	11210	11240	11270	11300	11330	11360	11390	11420	11450	11480	11510	11540	11570	11600	11630	11660	11690	11720	11750	11780	11810	11840	11870	11900	11930	11960	11990	12020	12050	12080	12110	12140	12170	12200	12230	12260	12290	12320	12350	12380	12410	12440	12470	12500	12530	12560	12590	12620	12650	12680	12710	12740	12770	12800	12830	12860	12890	12920	12950	12980	13010	13040	13070	13100	13130	13160	13190	13220	13250	13280	13310	13340	1337
----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	------

35 231 240 245 246 250 253 260  
 val ser lys leu val thr asp leu thr lys val his thr glu cys his gly asp leu leu glu cys ala asp asp arg ala asp leu  
 GTT TCC AAG TTA GTG ACA GAT CTT ACC AAA GTC CAC ACG GAA TGC TGC CAT GGA GAT CTG CTT GAA TGT GCT GAT GAC ACG CCG GAC CTT (890)  
  
 261 265 270 278 279 280 289 290  
 ala lys tyr ile cys glu asn gln ser ile ser ser lys leu lys glu cys cys glu lys pro leu leu glu lys ser his cys ile  
 GCC AAG TAT ATC TGT GAA AAT CAA GAT TCG ATC TCC AGT AAA CTG AAG GAA TGC TGT GGT GAA AAA CCT CTG TTG GAA AAA TCC CAC TGC ATT (980)  
  
 291 300 310 316 320  
 ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala  
 GCC GAA GTG GAA AAT GAT GAG ATG CCT GCT GAC TTG CCT TCA TTA GCT GCT GAT TTT GTT GAA AGT AAG CAT GTT TGC AAA AAC TAT GTT (1070)  
  
 321 330 340 350  
 glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu leu arg leu ala  
 GAG GCA AAG GAT GTC TTC TTG GCC ATG TTT TTG TAT GAA TAT GCA AGA AGG CAT CCT GAT TAC TCT GTC CTG CTG CTG ACA CTT GCC (1160)  
  
 351 360 361 369 370 380  
 lys thr tyr glu thr thr leu glu lys cys cys ala ala asp pro his glu cys tyr ala lys val phe asp glu phe lys pro leu  
 AAG ACA TAT GAA ACC ACT CTA GAG AAG TGC TGT GCC GCT GCT GCA GAT CCT CAT GAA TGC TAT GCC AAA GTG TTC GAT GAA TTT AAA CCT CCT (11250)  
  
 381 390 392 400 410  
 val glu glu pro gln asn leu ile lys gln asn cys glu leu phe glu gln leu qly glu tyr lys phe qln asn ala leu leu val arg  
 GTG GAA GAG CCT CAG AAT TTA ATC AAA CAA AAT TGT CAG CTT TTT CAG CAG CTT GCA GAG TAC AAA TTC CAG AAT CCG CTG TTA GTT CGT (11340)  
  
 411 420 430 437 438 440  
 tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu qly lys val qly ser lys cys cys lys his  
 TAC ACC AAG AAA GTA CCC CAA GTG TCA ACT CCA ACT CTT GTA CAG GTC TCA ACA AAC CTA GCA AAA GTG GGC AGC AAA TGT TGT AAA CAT (11430)  
  
 441 448 450 460 461 470  
 pro glu ala lys arg met pro cys ala glu asp tyr leu ser val val leu asn gln leu cys val leu his glu lys thr pro val ser  
 CCT GAA GCA AAA ACA ATG CCC TGT GCA GAA GAC TAT CTA TCC GTG GTC CTG AAC CAG TTA TGT GTG TTT CAT GAG AAA ACC CCA GTA AGT (11520)  
  
 471 476 477 490 500  
 asp arg val thr lys cys cys thr glu ser leu val asn arg pro cys phe ser ala leu glu val asp glu thr tyr val pro lys  
 GAC AGA GTC ACC AAA TCC TCC ACA GAA TCC TTG GTG AAC AGG CCA CCA TGC TTT TCA GCT CTG GAA GTC GAT GAA ACA TAC GTT CCC AAA (11610)  
  
 501 510 516 520 530  
 glu phe asn ala glu thr phe thr phe his ala asp ile cys thr leu ser glu lys glu arg qln ile lys lys qln thr ala leu val  
 GAG TTT AAT GCT GAA ACA TTC ACC TTC CAT GCA GAT ATA TCC ACA CTT TCT GAG AAG GAG AGA CAA ATC AAC AAA CAA ACT GCA CTT GTT (11700)



Following are examples which illustrate procedures, ~~including the best mode~~, for practicing the invention. These examples should not be construed as limiting. All percentages are by weight and all solvent mixture proportions are by volume unless otherwise noted.

5 Example 1      Isolation of Messenger RNA

Human liver mRNA was obtained following the procedure of Chirgwin, et al [Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299]. Immunoprecipitation of albumin containing polysomes was performed according to Taylor and  
10 Tse [Taylor, J.M. and Tse, T.P.H. (1976) J. Biol. Chem. 251, 7461-7467]. In vitro translation of mRNA was carried out in a reticulocyte cell-free system, following the instruction of the manufacturer (New England Nuclear). The translation products were separated electrophoretically according to Laemmli [Laemmli, J.K. (1970) Nature 227,  
15 680-685.

Example 2      Cloning Procedures

Double stranded cDNA was synthesized as described previously [Law, S., Tamaoki, T., Kreuzaler, F. and Dugaiczky, A. (1980) Gene 10, 53-61]. It was annealed to PstI-linearized pBR322 DNA [Rolivar, F.,  
20 Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crossa, J.H. and Falkow, S. (1977) Gene 2, 95-113] that had been tailed with 15 dG residues/3'-terminus [Dugaiczky, A., Robberson, D.L. and Ullrich, A. (1980) Biochemistry 19, 5869-5873]. The annealed DNA was used to transform E. coli strain RR1, as detailed previously [Law, S., et al., ibid.]. The albumin clones were selected using the colony  
25 hybridization method of Grunstein and Hogness [Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965], with [<sup>32</sup>P]-labeled cDNA synthesized with the immunoprecipitated polysomal mRNA as template.

30 As shown in Example 5, plasmids pHA36 and pHA206 were deposited in E. coli HB101 hosts. The plasmids were obtained from E. coli RR1 hosts, described in this example, and transformed into E. coli HB101 by standard procedures well known to those of ordinary skill in this art. The E. coli RR1 hosts were lysed and then centrifuged to  
35 separate the chromosomal DNA, cell DNA and plasmid DNA. The plasmid DNA, remaining in the supernatant, is precipitated with ethanol and the precipitate is resuspended in buffer, e.g., TCM (10mM Tris·HCl, pH 8.0, 10 mM CaCl<sub>2</sub>, 10 mM MgCl<sub>2</sub>). The cells for transformation are



prepared as follows: 120 ml of L-broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl) are inoculated with an 18 hour culture of HB101 NRRL B-11371 and grown to an optical density of 0.6 at 600 nm. Cells are washed in cold 100 mM NaCl and resuspended for 15 minutes in 20 ml  
5 chilled 50 mM CaCl<sub>2</sub>. Bacteria are then concentrated to one-tenth of this volume in CaCl<sub>2</sub> and mixed 2:1 (v:v) with annealed plasmid DNA, prepared as described above. After chilling the cell-DNA mixture for 15 minutes, it is heat shocked at 42°C for 2 minutes, then allowed to equilibrate at room temperature for ten minutes before addition of  
10 L-broth 10 times the volume of the cell-DNA suspension. Transformed cells are incubated in broth at 37°C for one hour before inoculating selective media (L-agar plus 10 µg/ml tetracycline) with 200 µl/plate. Plates are incubated at 37°C for 48 hours to allow the growth of transformants.

15 Example 3      Mapping of Restriction Endonuclease Sites

Restriction endonucleases were obtained from Bethesda Research Laboratories and New England Biolabs and were used according to the manufacturers' instructions. The digested DNA fragments were analyzed electrophoretically on agarose [Helling, R.B., Goodman, H.M. and  
20 Boyer, H.W. (1974) J. Virol. 14, 1235-1244] or acrylamide [Dingman, C., Fisher, M.P. and Kakefuda, T. (1972) Biochemistry 11, 1242-1250] gels.

Example 4      DNA Sequencing

DNA fragments were dephosphorylated with bacterial alkaline  
25 phosphatase (Worthington) and labeled at the 5'-ends with polynucleotide kinase (Boehringer-Mannheim) and  $\gamma$ [<sup>32</sup>P]ATP. Following digestion with a second restriction endonuclease and electrophoretic separation of the fragments, DNA sequence determination was done according to the procedure of Maxam and Gilbert [Maxam, A. and  
30 Gilbert, W. (1980) Methods Enzym. 65, 499-560] and the degradation products were separated electrophoretically on 0.4 mm acrylamide gels as described by Sanger and Coulson [Sanger, F. and Coulson, R. (1978) FEBS Letters 87, 107-110].

Example 5      Recombinant Plasmids pHA36 and pHA206

35 As disclosed in Example 2, albumin clones were selected by hybridizing to the enriched albumin cDNA probe. Plasmid pHA36 contained the largest insert of an albumin cDNA sequence. Both plasmids pHA36 and pHA206 have been deposited in a viable E. coli host in the

permanent collection of the Northern Regional Research Laboratory (NRRL), U.S. Department of Agriculture, Peoria, Illinois, U.S.A. Their accession numbers in this repository are as follows:

HB101(pHA36) - NRRL B-12551

5

HB101(pHA206) - NRRL B-12550

E. coli HB101 is a known and widely available host microbe. Its NRRL accession number is NRRL B-11371.

NRRL B-12550 and NRRL B-12551 are available to the public. ~~upon the grant of a patent. It should be understood that the availability~~  
10 of these deposits does not constitute a license to practice the subject invention in derogation of patent rights granted with the subject instrument by governmental action.

E. coli RR1 and E. coli HB101 are known and widely available host microbes. Their NRRL accession numbers are NRRL B-12186 and NRRL  
15 B-11371, respectively.

pBR322 is a well known and widely available plasmid. It can be obtained from the following host deposit by standard procedures:

NRRL B-12014 - E. coli RR1 (pBR322).

YEpl6 is a well known and widely available yeast episomal plasmid.  
20 It can be obtained from the following host deposit by standard procedures:

E. coli HB101 (YEpl6) - NRRL B-12093.

Example 6      Assembly of the Serum Albumin Gene

Assembling the pieces together is a straightforward task of re-  
25 striction enzymology. There is only one MspI site in the overlapping DNA sequence of the two cDNA clones. Two enzymatic steps of (i) MspI digestion of the two DNAs, followed by (ii) the use of ligase, an enzyme that seals DNA fragments, will give the desired product. Although two other undesired DNA species will also be obtained in the  
30 course of this recombination reaction, both of them will differ substantially in size. Thus, separation and isolation of the desired DNA species will be achieved.

The assembled DNA clone can be used to transform two types of cells:

35

(a) Escherichia coli

(b) Saccharomyces cerevisiae

(a) The vector of choice is plasmid pBR322, the same that has

been successfully used for cloning of the two fragmented pieces of the serum albumin cDNA.

(b) In order to transform yeast with the serum albumin structural gene sequence, the DNA must be inserted into one of the existing yeast plasmid vectors. This can be accomplished by taking advantage of the fact that several restriction endonuclease recognition sequences are absent from the cloned serum albumin DNA. Synthetic EcoRI DNA linkers can be ligated to the DNA fragment containing the serum albumin sequence followed by insertion (ligation) into one of the yeast plasmid vectors, e.g., YEp6, at the Eco RI cloning site. The fused chimeric plasmid can be used to transform yeast according to an established procedure [Hinnen, A., Hicks, J.R. and Fink, G.R. (1978) Proc. Natl. Acad. Sci. USA, 75, 1929]. YEp6 can be obtained from the NRRL repository, as disclosed supra.

#### 15 Example 7      Expression of the Serum Albumin Gene

The main body of the structural gene will be transcribed by the E. coli or yeast enzymes. If little or no albumin is produced with the selected host, then an Escherichia coli promoter DNA sequence carrying an initiation codon, i.e., ATG, can be ligated at the beginning of the serum albumin structural gene. Such elements are known and available, e.g., lac promoter used for the expression of human interferon gene in E. coli [Proc. Natl. Acad. Sci. 77, 5230 (1980)]; source of promoter DNA [Proc. Natl. Acad. Sci. 76, 760 (1979)]. Also, see Nature, Vol. 281, October 18, 1979. It has already been documented that such Escherichia coli promoter sequences function well in the expression of foreign genes in Escherichia coli [Mercereau-Puijalon, O., Royal, A., Cami, B., Garapin, A., Krust, A., Gannon, I. and Kourilsky, P. (1978) Nature 275, 505; and Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Grea, R., Hirose, T., Kraszewski, A., Itakura, K., and Riggs, A. (1979) Natl. Acad. Sci. USA 76, 106]. For expression in yeast, see Rose, M., Casadaban, M.J. and Botstein, D. (1981) Proc. Natl. Acad. Sci. USA 78, 2460 and 4466.

#### 30 Example 8      Screening of Clones Producing Albumin

Immunological methods can be used to detect small amounts of albumin made in a bacterium. Flat disks of flexible polyvinyl are coated with the IgG fraction from an immune serum and the disks are pressed onto an agar plate so that antigen released from an in situ lysed microbial colony can bind to the fixed antibody. The plastic

disk is then incubated with the same total IgG fraction labeled with radioactive iodine so that other determinants on the bound antigen can in turn bind the iodinated antibody. Radioactive areas on the disk expose X-ray film during autoradiography and thus identify colonies  
5 producing the protein which is being screened for. Detailed protocols of this procedure have been published [Broome, S. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA, 75, 2746]. The purification of human serum albumin can be accomplished by using procedures well known in the art. For example, procedures disclosed in a chapter by T.  
10 Peters: Purification and Properties of Serum Albumin, in: The Plasma Proteins, Putnam, Ed. Academic Press, New York, 1975, can be used.

The work described herein was all done in conformity with physical and biological containment requirements specified in the NIH Guidelines.

15

20

25

30

35

CLAIMS

1. Plasmid pHA36, having a restriction endonuclease pattern as shown in the drawing.

5

2. Plasmid pHA206, having a restriction endonuclease pattern as shown in the drawing.

3. E. coli HB101 (pHA36) having the deposit accession number  
10 NRRL B-12551.

4. E. coli HB101 (pHA206) having the deposit accession number  
NRRL B-12550.

15 5. A microorganism modified to contain a nucleotide sequence coding for the amino acid sequence of human serum albumin; said nucleotide sequence is as follows:

20

25

30

35

35  
 ser ala tyr ser arg gly val phe arg arg asp ala his lys ser glu val ala his arg phe lys asp leu gly glu asp phe lys  
 TCG GCT TAT TCC AGG GGT GTG TTT CGT CGA GAT GCA CAC AAG AGT GAG GTT GCT CAT CCG TTT AAA GAT TTC GCA GAA GAA AAT TTC AAA (170)  
 10  
 20  
 21  
 ala leu val leu ile ala phe ala gln tyr leu gln gln cys pro phe glu asp his val lys leu val asp glu val thr glu phe ala  
 GCC TTG GTG TTG ATT GCC TTT GCT CAG TAT CTT CAG CAG TGT CCA TTT CAA CAT CAT GTA AAA TTA GTG AAT GAA GTA ACT CAA TTT GCA (260)  
 40  
 50  
 51  
 53  
 lys thr cys val ala asp glu ser ala glu asp cys asp lys ser leu his thr leu phe gly asp lys leu cys thr val ala thr leu  
 AAA ACA TGT GTT GCT GAT GAG TCA GCT GAA AAT TGT GAC AAA TCA CTT CAT ACC CTT TTT CGA GAC AAA TTA TGC ACA GTT GCA ACT CTT (350)  
 70  
 75  
 80  
 81  
 arg glu thr tyr gly glu met ala asp cys cys ala lys gln glu pro gly arg asp glu cys phe leu gln his lys asp asp asp pro  
 CGT CAA ACC TAT GGT CAA ATG GCT GAC TGC TGT GCA AAA CAA GAA CCT GCG ACA AAT GAA TGC TTC TTG CAA CAC AAA GAT CAC AAC CCA (460)  
 100 101  
 110  
 111  
 asn leu pro arg leu val arg pro glu val asp val met cys thr ala phe his asp asp glu thr phe leu lys lys tyr leu try  
 AAC CTC CCC CCA TTG GTG ACA CCA GAG GTT GAT GTG ATG TGC ACT GCT TTT CAT GAC AAT GAA GAG ACA TTT TTG AAA AAA TAC TTA TAT (530)  
 120  
 124  
 130  
 140  
 141  
 glu ile ala arg arg his pro tyr phe tyr ala pro glu leu leu phe phe ala lys arg tyr lys ala ala phe thr glu cys cys gln  
 GAA ATT GCC ACA ACA CAT CCT TAC TTT TAT GCC CCG GAA CTC GAT GAA CTT CCG GAT CAA GCG AAG GCT TCG TCT GCC AAA CAG ACA CTC AAG TGT (620)  
 160  
 168 169 170  
 171  
 177  
 180  
 ala ala asp lys ala ala cys leu leu pro lys leu asp glu leu arg asp glu gly lys ala ser ser ala lys gln arg leu lys cys  
 GCT GCT CAT AAA GCT GCC TGC CTG TTG CCA AAG CTC GAT GAA CTT CCG GAT CAA GCG AAG GCT TCG TCT GCC AAA CAG ACA CTC AAG TGT (710)  
 190  
 200  
 201  
 ala ser leu gln lys phe gly glu arg ala phe lys ala trp ala val ala arg leu ser gln arg phe pro lys ala glu phe ala glu  
 GCC AGT CTC CAA AAA TTT GCA GAA ACA GCT TTC AAA CCA TGC CCA GTA GCT CCC CTG AGC CAG ACA TTT CCC AAA GCT GAG TTT GCA CAA (800)  
 210  
 220  
 230

-10  
 leu phe leu phe ser  
 CTT TTT CTC TTT ACC (30)

231 val ser lys leu val thr asp leu thr lys val his thr glu cys his gly asp leu leu glu cys ala asp asp arg ala asp leu 260  
 GTT TCC AAG TTA GTG ACA GAT CTT ACC AAA GTC CAC ACG GAA TGC TGC CAT GCA GAT CTG CTT GAA TGT GCT GAT GAC AGG GCG GAC CTT (890)  
 261 ala lys tyr ile cys glu asn gln asp ser ile ser ser lys leu lys glu cys glu lys pro leu leu glu lys ser his cys ile 289 290  
 CCC AAG TAT ATC TGT GAA AAT CAA GAT TCG ATC TCC AGT AAA CTG MAG GAA TCC TGT GAA AAA CCT CTG TTG GAA AAA TCC CAC TCC ATT (980)  
 291 ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala 320  
 GCC CAA GTG GAA AAT GAT GAG ATG CCT GCT GAT TTA GCT CCT TCA TTA GCT GCT GAT TTT GTT GAA AGT AAG GAT GTT TGC AAA AAC TAT GCT (1070)  
 321 glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu leu arg leu ala 350  
 GAG GCA AAG GAT GTC TTC TTG GCG ATG TTT TTG TAT GAA TAT GCA AGA AGG CAT CCT GAT TAC TCT GTC GTG CTG CTG ACA CTT GCC (1160)  
 351 lys thr tyr glu thr thr leu glu lys cys ala ala asp pro his glu cys tyr ala lys val phe asp glu phe lys pro leu 380  
 AAG ACA TAT GAA ACC ACT CTA GAG AAG TGC TGT GCT GCT GCA GAT CCT CAT GAA TGC TAT GCC AAA GTG TTC GAT GAA TTT AAA CCT CCT (1250)  
 381 val glu glu pro gln asn leu ile lys gln asn cys glu leu phe glu qln leu qly glu tyr lys phe qln asn ala leu leu val arg 410  
 GTG GAA GAG CCT CAG AAT TTA ATC AAA CAA AAT TGT GAG CAG CTT GCA GAG TAC AAA TTC CAG AAT CCG CTG TTA GTT CGT (1340)  
 411 tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu qly lys val qly ser lys cys cys lys his 440  
 TAC ACC AAG AAA GTA CCC CAA GTG TCA ACT CCA ACT CTT GTA GAG GTC TCA AGA AAC CTA GCA AAA GTG GCC AGC AAA TGT TGT AAA CAT (1430)  
 441 pro glu ala lys arg met pro cys ala glu asp tyr leu ser val val leu asn gln leu cys val leu his qly lys thr pro val ser 470  
 CCT GAA GCA AAA ACA ATG CCC TGT GCA GAA GAC TAT CTA TCC GTG GTC CTG AAC CAG TTA TGT GTG TTA TGT CAT GAG AAA ACG CCA GTA AGT (1520)  
 471 asp arg val thr lys cys cys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val asp glu thr tyr val pro lys 500  
 CAC ACA GTG ACC AAA TGC TCC ACA GAA TCC TTG GTG AAC AGG CGA CCA TCC TTT TCA GCT CTG GAA GTC GAT GAA ACA TAC GTT CCC AAA (1610)  
 501 glu phe asn ala glu thr phe thr phe his ala asp ile cys thr leu ser glu lys glu arg qln ile lys lys qln thr ala leu val 530  
 GAG TTT AAT GCT GAA ACA TTC ACC TTC CAT GCA GAT ATA TCC ACA CTT TCT CAG AAG GAG AGA CAA ATC AAG AAA CAA ACT GCA CTT GTT (1700)

5  
 10  
 15  
 20  
 25  
 30  
 35

531  
 glu leu val lys his lys pro lys ala thr lys glu gln leu lys ala val met asp phe ala ala phe val glu lys cys lys  
 GAG CTC GTG AAA CAC AAG CCC AAG GCA ACA AAA GAG CAA CTG AAA GCT GTT ATG CAT GAT TTC GCT TTT GTA GAG AAG TGC AAG (1790)

540  
 550  
 558 559 560

561  
 ala asp asp lys glu thr cys phe ala glu glu gln lys lys leu val ala ala ser gln ala ala leu gly leu ter  
 GCT CAC CAT AAG GAG ACC TGC TTT GCC GAG GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GCC TTA TAA CATCACATTTAAAG (1883)

ter ter

CATCTCAGCCTACCATGAGATATAGACAGAAAGAAATGAAATGAAAGCTTATTCATCTGTTTTTTCGTTGGTGAAGCCACACCCCTGCTATAAAACATAAATTTCTTTAA (2002)

TCATTTTGGCTCTCTTTCTCTGCTTCAATTAATAAAATGAAAGAAATCTAA..... 20 .....AA (2078)



6. Nucleotide sequence of the cDNA of human serum albumin, said nucleotide sequence is as follows:

```

5
10
15
20
25
30
35
1      10      20
asp ala his lys ser glu val ala ala his arg phe lys asp leu gly glu glu asn phe lys
CAT CCA CAC AAG AGT CAG GTT GCT CAT CCG TTT AAA GAT TTG CCA GAA GAA AAT TTC AAA (170)

21      30      40      50
ala leu val leu ile ala phe ala gln tyr leu gln gln cys pro phe glu asp his val lys leu val asn glu val thr glu phe ala
GCC TTG GTG TTG ATT GCC TTT GCT CAG TAT CTT CAG CAG TGT CCA TTT GAA GAT CAT GTA AAA TTA GTG AAT GAA GTA ACT CAA TTT CCA (260)

51      60      70      75      80
lys thr cys val ala asp glu ser ala glu asn cys asp lys ser leu his thr leu phe gly asp lys leu cys thr val ala thr leu
AAA ACA TGT GTT GCT GAT CAG TCA GCT GAA AAT TGT CAG AAA TCA CTT CAT ACC CTT TTT CGA CAC AAA TTA TCC ACA GTT CCA ACT CTT (350)

81      90      91      100 101      110
arg glu thr tyr gly glu met ala asp cys cys ala lys gln glu pro gly arg asn glu cys phe leu gln his lys asp asp asn pro
CGT GAA ACC TAT GGT GAA ATG CCT CAC TGC TGT CCA AAA CAA CCA CCT GGG AGA AAT CAA TGC TTC TTG CAA CAC AAA GAT GAC AAC CCA (440)

111      120      124      130      140
asn leu pro arg leu val arg pro glu val asp val met cys thr ala phe his asp asn glu thr phe leu lys lys tyr leu try
AAC CTC CCC CCA TTG GTG ACA CCA CAG GTT GAT GTG ATG TGC ACT GCT TTT CAT GAC AAT GAA CAG ACA TTT TTG AAA AAA TAC TTA TAT (530)

141      150      160      168 169 170
glu ile ala arg arg his pro tyr phe phe ala pro glu leu leu phe phe ala lys arg tyr lys ala ala phe thr glu cys cys gln
CAA ATT GCC ACA ACA CAT CCT TAC TTT TAT GCC CCG GAA CTC CTT TTC TTT GCT AAA AGG TAT AAA GCT GCT TTT ACA CAA TGT TGC CAA (620)

171      177      180      190      200
ala ala asp lys ala ala cys leu leu pro lys leu asp glu leu arg asp glu gly lys ala ser ser ala lys gln arg leu lys cys
CCT GCT CAT AAA GCT GCC TGC CTG TTG CCA AAG CTC GAT GAA CTT CCG CAT GAA GGG AAG GCT TCG TCG GCC AAA CAG ACA CTC AAG TGT (710)

201      210      220      230
ala ser leu gln lys phe gly glu arg ala phe lys ala trp ala val ala arg leu ser gln arg phe pro lys ala glu phe ala glu
GCC AGT CTC CAA AAA TTT GCA GAA AGA GCT TTC AAA CCA TGG GCA GTA GCT CCC CTG AGC CAG ACA TTT CCC AAA GCT CAG TTT GCA GAA (800)

```

231	240	245 246	250	253	260
val ser lys leu val thr asp leu thr lys val his thr glu cys cys his gly asp leu leu glu cys ala asp asp arg ala asp leu					
GTT TCC AAG TTA GTG ACA CAT CTT ACC AAA GTC CAC ACG GAA TCC TGC CAT GCA CAT CTG CTT GAA TGT CCT GAT CAC ACG CCG CAC CTT (890)					
261	270	278 279 280			289 290
ala lys tyr ile cys glu asn gln asp ser ile ser ser lys leu lys glu cys cys glu lys pro leu leu glu lys ser his cys ile					
CCC AAG TAT ATC TGT GAA AAT CAA CAT TCG ATC TCC AGT AAA CTG AAG GAA TCC TGT GAA AAA CCT CTG TTG GAA AAA TCC CAC TCC ATT (980)					
291	300	310	316		320
ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala					
CCC GAA GTG GAA AAT CAT CAG ATG CCT GCT CAC TTG CCT TCA TTA CCT CCT GAT TTT CTT CAA AGT AAG CAT GTT TGC AAA AAC TAT CTT (1070)					
321	330	340			350
glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu leu arg leu ala					
CAG CCA AAG CAT GTC TTC TTG GCC ATG TTT TTG TAT GAA TAT GCA AGA AGG CAT CCT CAT TAC TCT GTC GTG CTG CTG AGA CTT GCC (1160)					
351	360 361	369 370			380
lys thr tyr glu thr thr leu glu lys cys cys ala ala asp pro his glu cys tyr ala lys val phe asp glu phe lys pro leu					
AAG ACA TAT GAA ACC ACT CTA CAG AAG TCC TGT GCC CCT GCA CAT CCT CAT GAA TCC TAT GCC AAA GTG TTC GAT GAA TTT AAA CCT CCT (1250)					
381	390 392	400			410
val glu glu pro gln asn leu ile lys gln asn cys glu leu phe glu qln leu gly tyr lys phe qln asn ala leu leu val arg					
GTG GAA CAG CCT CAG AAT TTA ATC AAA CAA AAT TGT CAG CTT TTT CAG CAG CTT GCA CAG TAC AAA TTC CAG AAT CCG CTG TTA GTT CCT (1360)					
411	420	430		437 438	440
tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu gly lys val qly ser lys cys cys lys his					
TAC ACC AAG AAA GTA CCC CAA GTG TCA ACT CCA ACT CTT GTA CAG GTC TCA AGA AAC CTA GCA AAA GTG GCC AGC AAA TGT TGT AAA CAT (1430)					
441	448 450	460 461			470
pro glu ala lys arg met pro oys ala glu asp tyr leu ser val val leu asn gln leu cys val leu his qlu lys thr pro val ser					
CCT GAA GCA AAA ACA ATG CCC TGT GCA GAA CAG TAT CTA TCC GTG GTC CTG AAC CAG TTA TGT GTG TTG CAT GAG AAA ACG CCA GTA ACT (1520)					
471	476 477	490			500
asp arg val thr lys cys cys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val asp qlu thr tyr val pro lys					
GAC AGA GTC ACC AAA TGC TCC ACA GAA TCC TTG GTG AAC AGG CCA CCA TCC TTT TCA CCT CTG GAA GTC CAT GAA ACA TAC GTT CCC AAA (1610)					
501	510	514	520		530
glu phe asn ala glu thr phe thr phe his ala asp ile cys thr leu ser glu lys glu arg qln ile lys lys qln thr ala leu val					
CAG TTT AAT CCT GAA ACA TTC ACC TTC CAT CCA GAT ATA TCC ACA CTT TCT CAG AAG GAG ACA CAA ATC AAG AAA CAA ACT CCA CTT GTT (1700)					

5  
10  
15  
20  
25  
30  
35

531  
glu leu val lys his lys pro lys ala thr lys glu gln leu lys ala val met asp phe ala ala phe val glu lys cys lys  
GAG CTC GTG AAA CAC CAC AAG CCC AAG CCA ACA AAA GAG CAA CTG AAA GCT GTT ATG CAT CAT TTC GCT GCT TTT GTA GAG AAG TGC AAG (1790)

540  
550  
558 559 560  
561  
ala asp asp lys glu thr cys phe ala glu gln lys leu val ala ala ser gln ala ala leu gly leu ter  
GCT GAC GAT AAG GAG ACC TGC TTT GCC GAG GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA TAA CATCACATTTAAAG (1883)

567  
570  
580  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000

5

10

15

20

25

30

35

Met lys trp val tlu phe ile ser leu leu phe leu phe ser  
ATG AAG TGG GTA ACC TTT ATT TCC CTT TTT CTC TTT ACC (30)

GCCTTTCTCTCTCTGTCACCCACAGCCCTTTGGCACA

ser ala tyr ser arg gly val phe arg arg  
TCG GCT TAT TCC ACC GGT GTG TTT CGT CGA

35	-6	p	r	o	-1	1	10	20	25	30	35	5																				
	arg	gly	val	phe	arg	asp	ala	his	lys	ser	glu	val	ala	his	arg	phe	lys	asp	leu	gly	glu	asn	phe	lys	20							
	ACG	GGT	GTG	TTT	CGT	CCA	GAT	GCA	CAC	AGC	AGT	CAG	GTT	GCT	CAT	CGG	TTT	AAA	GAT	TTG	GCA	GAA	AAAT	TTC	AAA	(170)						
21	ala	leu	val	leu	ile	ala	phe	ala	gln	tyr	leu	gln	gln	cys	pro	phe	glu	asp	his	val	lys	leu	val	asn	glu	val	thr	glu	phe	ala	50	
	CCC	TTG	GTG	TTG	ATT	GCC	TTT	GCT	CAG	TAT	CTT	CAG	CAG	TGT	CCA	TTT	GAA	CAT	CAT	GTA	AAA	TTA	GTG	AAT	GAA	GTA	ACT	CAA	TTT	GCA	(260)	
51	lys	thr	cys	val	ala	asp	glu	ser	ala	glu	asn	cys	asp	lys	ser	leu	his	thr	phe	gly	asp	lys	leu	cys	thr	val	ala	thr	leu	80		
	AAA	ACA	TGT	GTT	GCT	CAT	CAG	TCA	GCT	GAA	AAT	TGT	GAC	AAA	TCA	CTT	CAT	ACC	CTT	TTT	CGA	GAC	AAA	TTA	TGC	ACA	GTT	GCA	ACT	CTT	(350)	
81	arg	glu	thr	tyr	gly	glu	met	ala	asp	cys	cys	ala	lys	gln	glu	pro	gly	arg	asn	glu	cys	phe	leu	gln	his	lys	asp	asp	asn	pro	110	
	CGT	GAA	ACC	TAT	GGT	GAA	ATG	CCT	GAC	TGC	TGT	GCA	AAA	CAA	GAA	CCT	GGG	ACA	AAT	CAA	TGC	TTT	TTG	CAA	CAC	AAA	GAT	GAC	AAC	CCA	(660)	
111	asn	leu	pro	arg	leu	val	arg	pro	glu	val	asp	val	met	cys	thr	ala	phe	his	asp	asn	glu	glu	thr	phe	leu	lys	lys	tyr	leu	try	140	
	AAC	CTC	CCC	CCA	TTG	GTG	ACA	CCA	GAG	GTT	GAT	GTG	ATG	TGC	ACT	GCT	TTT	CAT	GAC	AAT	GAA	GAG	ACA	TTT	TTG	AAA	AAA	TAC	TTA	TAT	(330)	
141	glu	ile	ala	arg	arg	his	pro	tyr	phe	tyr	ala	pro	glu	leu	leu	phe	phe	ala	lys	arg	tyr	lys	ala	ala	phe	thr	glu	cys	cys	gln	160	
	GAA	ATT	CCC	ACA	ACA	CAT	CCT	TAC	TTT	TAT	CCC	CCG	GAA	CTC	CTT	TTT	GCT	AAA	AGG	TAT	AAA	GCT	GCT	TTT	ACA	CAA	TGT	TGC	CAA	(620)		
171	ala	ala	asp	lys	ala	ala	cys	leu	leu	pro	lys	leu	asp	glu	leu	arg	asp	glu	gly	lys	ala	ser	ser	ala	lys	gln	arg	leu	lys	cys	200	
	GGT	GCT	GAT	AAA	GCT	CCC	TGC	TGC	TTG	CTG	CTG	CTG	CTG	GAT	GAA	CTT	CCG	GAT	CAA	GGG	AAG	GCT	TGC	TCT	CCC	AAA	CAG	ACA	CTC	AAG	TGT	(710)
201	ala	ala	ser	leu	gln	lys	phe	gly	glu	arg	ala	phe	lys	ala	trp	ala	val	ala	arg	leu	ser	gln	arg	phe	pro	lys	ala	glu	phe	ala	glu	230
	GGC	AGT	CTC	CAA	AAA	TTT	GCA	GAA	AGA	GCT	TTC	AAA	GCA	TGG	GCA	GTA	GCT	CCC	CTG	AGC	CAG	AGA	TTT	CCC	AAA	GCT	GAG	TTT	GCA	CAA	(300)	

231 val ser lys leu val thr asp leu thr lys val his thr glu cys cys his gly asp leu leu glu cys ala asp asp arg ala asp leu  
 GTT TCC AAG TTA GTG ACA GAT CTT ACC AAA GTC CAC ACG GAA TGC TCC CAT GCA GAT CTG CTT GAA TGT GCT CAT CAC ACG GCG GAC CTT (890)

261 ala lys tyr ile cys glu asn gln asp ser ile ser ser lys leu lys leu lys glu cys glu lys pro leu leu glu lys ser his cys ile  
 GCC AAG TAT ATC TGT GAA AAT CAA GAT TCG ATC TCC AGT AAA CTG AAG GAA TGC TGT GAA AAA CCT CTG TTC GAA AAA TCC CAC TCC ATT (980)

291 ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala  
 GCC GAA GTG GAA AAT GAT CAG ATG CCT CCT GAT TTA CCT GCT CAT TTT GTT GAA AGT AAG GAT GTT TCC AAA AAC TAT GCT (1070)

321 glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu arg leu ala  
 GAG GCA AAG GAT GTC TTC TTG GCC ATG TTT TTG TAT GAA TAT GCA AGA ACG CAT CCT GAT TAC TCT GTC GTG CTG CTA GCA CTT GCC (1160)

351 lys thr tyr glu thr thr leu glu lys cys cys ala ala asp pro his glu cys tyr ala lys val phe asp glu phe lys pro leu  
 AAG ACA TAT GAA ACC ACT CTA GAG AAG TGC TGT GCT GCA GAT CCT CAT GAA TGC TAT CCC AAA GTG TTC CAT GAA TTT AAA CCT CTT (1250)

381 val glu glu pro gln asn leu ile lys gln asn cys glu leu phe glu thr lys phe gln asn ala leu leu val arg  
 GTG GAA GAG CCT CAG AAT TTA ATC AAA CAA AAT TGT GAG CAG CTT TTT GAG CAG CTT GCA GAG TAC AAA TTC CAG AAT GCG CTG TTA GTT CTT (1340)

411 tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu qly lys val qly ser lys cys cys lys his  
 TAC ACC AAG AAA GTA CCC CAA GTG TCA ACT CCA ACT CTT GTA CAG GTC TCA ACA AAC CTA GCA AAA GTG GCG ACG AAA TGT TGT AAA CAT (1430)

441 pro glu ala lys arg met pro cys ala glu asp tyr leu ser val val leu asn gln leu cys val leu his glu lys thr pro val ser  
 CCT GAA GCA AAA AGA ATG CCC TGT GCA GAA GAC TAT CTA TCC GTC GTC AAC CAG TTA TGT GTC TTC CAT GAG AAA ACG CCA GTA AGT (1520)

471 asp arg val thr lys cys cys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val asp glu thr tyr val pro lys  
 GAC AGA GTC ACC AAA TCC TCC ACA GAA TCC TTG GTG AAC AGG GCA CCA TCC TTT TCA GCT CTG CAA GTC CAT GAA ACA TAC GTT CCC AAA (1610)

501 glu phe asn ala glu thr phe thr phe his ala asp ile cys thr leu ser glu lys glu arg gln ile lys lys gln thr ala leu val  
 GAG TTT AAT GCT GAA ACA TTC ACC TTC CAT GCA GAT ATA TCC ACA CTT TCT GAG AAG AAG ACA ATC AAG AAA CAA ACT GCA CTT GTT (1700)

35

	-18	p r c	-10	
	Met lys trp val tlu phe lle ser leu leu phe leu phe ser ATC AAC TCG GTA ACC TTT ATT TCC CTT CTT TTT CTC TTT AGC (30)			
		10	20	
	ser ala tyr ser arg gly val phe arg arg asp ala his lys ser glu val ala his arg phe lys asp leu gly qlu qlu asn phe lys TCG GCT TAT TCC ACG GGT GTG TTT CGT CCA CAT CCA CAC AAC AGT CAG GTT GCT CAT CGG TTT AAA GAT TTG GCA GAA GAA AAT TTC AAA (170)			
21		30	40	50
ala leu val leu lle ala phe ala gin tyr leu gln gln cys pro phe qlu asp his val lys leu val asn glu val thr qlu phe ala GCC TTG GTG TTG ATT GCC TTT GCT CAG TAT CTT CAG CAG TGT CCA TTT GAA CAT CAT GTA AAA TTA GTC AAT GAA GTA ACT CAA TTT GCA (260)				
51	53	60	70	80
lys thr cys val ala asp glu ser ala glu asn cys asp lys ser leu his thr leu phe gly asp lys leu cys thr val ala thr leu AAA ACA TGT GTT GCT CAT GAT CAG TCA GCT GAA AAT TGT GAC AAA TCA CTT CAT ACC CTT TTT CCA CAC AAA TTA TCC ACA GTT GCA ACT CTT (350)				
81		90	100 101	110
arg glu thr tyr gly glu met ala asp cys cys ala lys gln glu pro gly arg asn qlu cys phe leu qln his lys asp asp asn pro CGT GAA ACC TAT GGT GAA ATG CAC TGC TGT GCA AAA CAA GAA CCT GGG ACA AAT GAA TGC TTC TTG CAA CAC AAA GAT GAC AAC CCA (660)				
111		120	130	140
asn leu pro arg leu val arg pro glu val asp val met cys thr ala phe his asp asn qlu qlu thr phe leu lys lys tyr leu try AAC CTC CCC CCA TTG GTG ACA CCA CAG GTT GAT GTG ATG TGC ACT GCT TTT CAT CAC AAT GAA GAG ACA TTT TTG AAA AAA TAC TTA TAT (730)				
141		150	160	170
glu lle ala arg arg his pro tyr phe tyr ala pro glu leu leu phe phe ala lys arg tyr lys ala ala phe thr qlu cys cys qln GAA ATT GCC ACA ACA CAT CCT TAC TTT TAT GCC CCG GAA CTC CTT TTC TTT GCT AAA AGG TAT AAA GCT GCT TTT ACA GAA TGT TGC CAA (620)				
171		180	190	200
ala ala asp lys ala ala cys leu leu pro lys leu asp glu leu arg asp qlu gly lys ala ser ser ala lys qln arg leu lys cys GCT GCT GAT AAA GCT GCC TGC CTG TTG CCA AAG CTC GAT GAA CTT CGG CAT GAA GGG AAG GCT TCG TCT GCC AAA CAG ACA CTC AAG TGT (710)				
201		210	220	230
ala ser leu gln lys phe gly glu arg ala phe lys ala trp ala val ala arg leu ser gln arg phe pro lys ala qlu phe ala qlu GCC AGT CTC CAA AAA TTT GGA GAA AGA GCT TTC AAA GCA TGG GCA GTA GCT CGC CTC AGC CAG ACA TTT CCC AAA GCT CAG TTT GCA CAA (300)				



231 val ser lys leu val thr asp leu thr lys val his thr glu cys oys his gly asp leu leu glu cys ala asp asp arg ala asp leu 260  
 GGT TCC AAG TTA GTG ACA GAT CTT ACC AAA GTC CAC ACC GAA TCC TCC CAT GCA GAT CTG CTT GAA TGT CAT CAT CAC AGG CCG CAC CTT (R90)  
  
 261 ala lys tyr lle cys glu asn gln asp ser lle ser ser lys leu lys leu lys pro leu leu glu lys ser his cys lle 289 290  
 CCC AAG TAT ATC TGT GAA AAT CAA GAT TCG ATC TCC AGT AAA CTG AAG CAA TCC TGT CAA AAA CCT CTG TTG CAA AAA TCC CAC TCC ATT (R90)  
  
 291 ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala 320  
 GCC GAA GTG GAA AAT GAT CAG ATG CCT GCT CAC TTG CCT TCA TTA GCT GCT CAT TTT GTT GAA AGT AAG CAT GTT TCC AAA AAC TAT GCT (1070)  
  
 321 glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu leu arg leu ala 350  
 GAG GCA AAG GAT GTC TTC TTG GGC ATG TTT TTG TAT GAA TAT GCA ACA AGG CAT CCT CAT TAC TCT GTC GTG CTG CAG CTT GCC CTT GCC (1160)  
  
 351 lys thr tyr glu thr leu glu lys cys ala ala asp pro his glu cys tyr ala lys val phe asp glu phe lys pro leu 380  
 AAG ACA TAT GAA ACC ACT CTA GAG AAG TCC TGT GCC GCT GCA CAT CCT CAT GAA TGC TAT GCC AAA GTG TTC GAT GAA TTT AAA CCT CCT (1250)  
  
 381 val glu glu pro gln asn leu lle lys gln asn cys glu leu phe glu leu phe gln asn ala leu leu val arg 410  
 GTG GAA CAG CCT CAG AAT TTA ATC AAA CAA AAT TGT GAG CAG CTT TTT GAG CAG CTT GGA GAG TAC AAA TTC CAG AAT CCG CTG TTA GTT CGT (1340)  
  
 411 tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu qly lys val qly ser lys cys cys lys his 440  
 TAC ACC AAG AAA GTA CCC CAA GTG TCA ACT CCA ACT CTT GTA GAG GTC TCA ACA AAC CTA GGA AAA GTG GCC ACC AAA TGT TGT AAA CAT (1430)  
  
 441 pro glu ala lys arg met pro cys ala glu asp tyr leu ser val val leu asn gln leu cys val leu his glu lys thr pro val ser 470  
 CCT GAA GCA AAA ACA ATG CCC TGT GCA GAA GAC TAT CTA TCC GTG GTC CTG AAC CAG TTA TGT GTG TTG CAT GAG AAA ACC CCA GTA AGT (1520)  
  
 471 asp arg val thr lys cys cys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val asp glu thr tyr val pro lys 500  
 CAC ACA GTC ACC AAA TCC TCC ACA GAA TCC TTG GTG AAC AGG CCA CCA TCC TTT TCA CCT CTG GAA GTC GAT CAA ACA TAC GTT CCC AAA (1610)  
  
 501 glu phe asn ala glu thr phe thr his ala asp lle cys thr leu ser glu lys glu arg gln lle lys lys gln thr ala leu val 530  
 GAG TTT AAT GCT GAA ACA TTC ACC TTC CAT GCA GAT ATA TCC ACA CTT TCT CAG AAG GAG AGA CAA ATC AAG AAA CAA ACT CCA CTT GTT (1700)

0079739

4083

-25-

5  
10  
15  
20  
25  
30  
35

531  
glu leu val lys his lys pro lys ala thr lys glu gln leu lys ala val met asp phe ala ala phe val glu lys cys lys  
CAG CTC GTG AAA CAC CAG CCC AAG CCA ACA CAG CAA CTC AAA GCT GTT ATG GAT CAT TTC GCT TTT GTA CAG AAG TGC AAG (1790)

540  
558 559 560  
561  
ala asp asp lys glu thr cys phe ala glu glu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter  
CCT GAC GAT AAG CAG ACC TGC TTT GCC CAG CAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA TAA CATCACATTAAAG (1883)

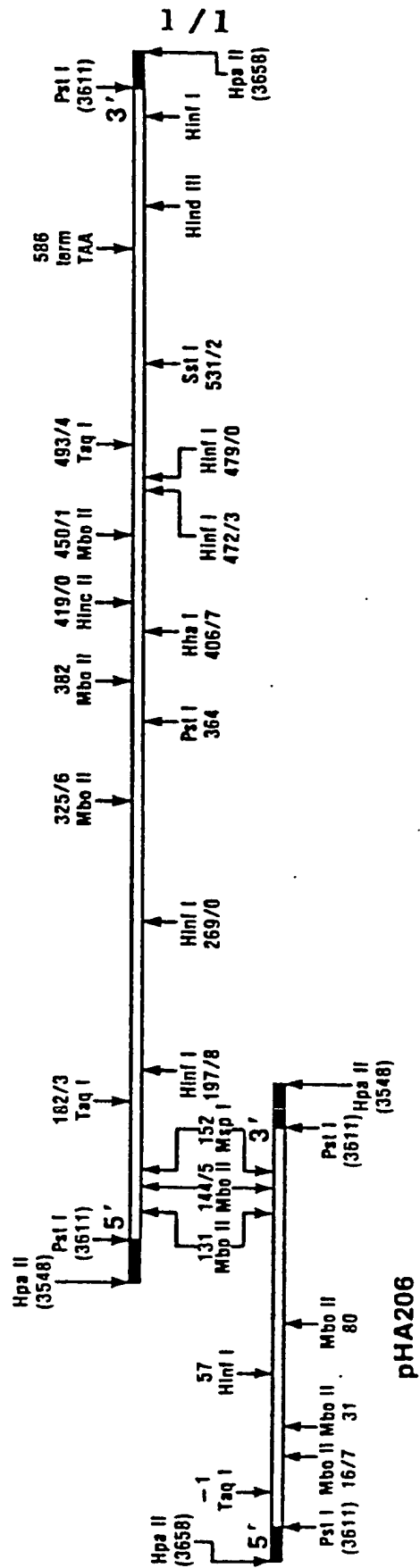
567  
570  
580  
ter ter  
CATCTCAGCCTACCATGAGATAAGAGAGAAATGAGATCAAAAGCTTATTCATCTGTTTTCTTTTCGTTGGTGTAAAGCCACCCCTGCTTAAACATATAATTTCTTTAA (2002)

TCATTTTGGCCTCTTTTCTCTGCTGCTTCAATTAAATAAAAAATGCAAGCAATCTAA..... 20 .....AA (2078)

10. A nucleotide sequence according to any of claims 6 to 9, in essentially pure form.
11. A DNA transfer vector comprising a nucleotide sequence as defined in claim 5.
- 5 12. A DNA transfer vector according to claim 11, transferred to and replicated in a micro-organism.
13. A DNA transfer vector according to claim 12, which is a plasmid.
14. A DNA transfer vector according to claim 13,
- 10 wherein the plasmid is pBR322 or YEp6.
15. A process for preparing human serum albumin, which comprises culturing a micro-organism according to claim 5.
16. A DNA transfer vector according to any of
- 15 claims 12 to 14, or a process according to claim 15, wherein the micro-organism is a bacterium or yeast.
17. A vector or process according to claim 16, wherein the bacterium or yeast is E. coli or Saccharomyces cerevisiae.

# Restriction Endonuclease Map of Human Serum Albumin cDNA Clones

pHA36



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**